

Using Digital Technologies is Securing Private and Public Interests in Gene Research

Sergey Borodin^{1, *} Valentina Ruzanova¹

¹ Samara National Research University, Samara 443086, Russian Federation

*Corresponding author. Email: borodinss@lenta.ru

ABSTRACT

The article is devoted to the definition of the specifics of the formation of intellectual property rights in the implementation of genomic research using technologies for the mining of big data, including in the field of circadian genetics, while respecting private and public interests. The author analyzes the regulation of the transfer and processing of personal data, including in the field of circadian genetics, the creation of new results of intellectual activity, taking into account the different approaches of scientists in the relevant fields, and law enforcement. It is proposed to legislatively provide for the power to regulate data placement, use of virtual computers and data management in the cloud, as well as fix the corresponding minimum requirements at the level of international regulation. It is concluded that, in terms of cross-border data transfer, it is advisable to use a comprehensive model of "adequate protection", "safe harbor" and "accountability" in the exchange of information between research institutes, universities and companies. It is indicated that at present, judicial practice is being formed to consider disputes related to research and certain manipulations with genetic material precisely from the position of a balance between public interests (morality) and private non-property interests (respect for private and family life), regardless of the assessment of any property interests in relation to the relevant objects. It is noted that, in the framework of further discussion, it is advisable to consider the development of a comprehensive legal regime for non-property rights to genes (including circadian genes) of a person in combination with a sui generis regime of intellectual rights. The article is of practical value not only for lawyers, but also for specialists in other fields (biology, medicine, etc.) in terms of their use of genomic information. The conclusions made by the authors are also applicable in the process of implementing legislative activity.

Keywords: *Digital technologies, gene research, private and public interests, legal provisions, ethics.*

1. INTRODUCTION

Currently, the collection and processing of Big Data is becoming increasingly important for the development of all spheres of society, including scientific research and the application of its results. As of 2014, 2.5 quintillion (10¹⁸) bytes of data were generated per day, 90 percent of the data in the modern world was created in just two years, while this data is characterized by a significant increase in volume, speed, variety and reliability [1]. It should be especially noted that over the past 8 years, the volume of genomic information has doubled every 5 months. However, despite the fact that the human genome was sequenced back in 2000, only 10-15% of genes have been identified so far, the function of 99% of DNA remains unknown [2].

A significant quantitative increase in the volume and qualitative change in the characteristics of data, in combination with the significantly increased computing capabilities of computers, as well as the introduction of cloud computing technologies, data mining, machine learning, has created new opportunities for carrying out genetic research (manipulation) with genes, including circadian. At the same time, scientists and potential users faced a number of serious challenges, including

not only the issues of legal identification of the results obtained and the establishment of the most appropriate regime for their protection, but also the problems of maintaining the privacy of research participants, including clinical ones, as well as ethical issues.

For example, a survey of more than 80,000 health-related websites found that in nine out of ten cases, personal health information was leaked to third parties, including advertisers and data brokers [1]. In the literature in relation to research projects, it is noted that most of the existing solutions that allow you to safely carry out data mining entail significant costs of computing power when implementing large-scale and complex tasks typical for biomedicine. This resource constraint forces researchers to rely on small datasets or simplified privacy practices [3]. Thus, research medical institutions collect a significant amount of information about the patients in them, not only for treatment, but also to replenish their own data banks [4]. Thus, in modern conditions, the problem often arises of choosing between ensuring the completeness and reliability of the conclusions obtained as a result of research, and creating conditions for the proper provision of the privacy of persons whose data are used.

In addition, the transfer of genomic and clinical data is faced with a regulation originally designed to protect

personal data only within one jurisdiction. Such data protection systems do not meet the needs of the growing large-scale, world-class health research, which often leads to inefficiencies in data exchange, significant costs and delays in information processing [5].

Note that genomic research is one of the key areas for the development of the digital economy, and therefore the priority task is to identify the optimal ways to include their results in the civil turnover [6], [7]. Therefore, in addition to the problems of ensuring privacy when processing personal information, there is also the question of establishing an appropriate regime for the legal protection of the results of intellectual activity. An important role in this is acquired by the aspect of compliance of such studies (manipulations) with the principles of humanity and morality, which, in turn, is reflected in the legislation. For example, for the first time in the practice of European supranational regulation, the need to comply with ethical standards as part of the legality of data processing in the framework of research was referred to in the Regulation of the European Parliament and the Council of the European Union 2016/679 of 04/27/2016 (hereinafter - Regulation 2016/679) [8].

This work contains the results of consideration of the entire process of creating and formalizing new results of intellectual activity, set out in a systematic form, ranging from the use of technologies for data mining and cross-border transfer of genetic data, confidentiality and respect for private life, ensuring public interests, and up to possible options for establishing legal protection and gene sequences, as well as to assess the prospects for the use of information technology in the implementation of scientific research, including in the field of circadian genetics.

2. ENSURING THE RIGHT TO PRIVACY

Currently, a new area of experimental research is being built - *in silico* biology, which, in contrast to *in vivo* (in vivo, inside a living organism) and *in vitro* (in vitro), studies biological phenomena and aspects based on computer modeling. Such technologies, used in the development and testing of new drugs and treatments, save time and money, as well as, most importantly, the lives of millions of animals [2]. In turn, the implementation of *in silico* experiments requires the transfer of genetic data for processing, including through cloud computing, as well as the formation of decision-making techniques within the framework of the corresponding data mining.

The main result of the application of these technologies is a clinically useful classification of patients, constructed on the basis of assumptions about which

data elements are significant in terms of distinguishing phenotypic groups [9]. The fundamental, therefore, is the search for patterns in the gigantic volume of collected data, including the relationships previously missed by the human researcher. This information creates the preconditions for the development of personalized medicine, since it not only makes it possible to diagnose a disease, but also to identify mutations that are most effectively amenable to drug treatment [4]. Thus, the conclusions obtained as a result of the mining of big data can be useful not only within the framework of a specific task set during a certain study, but also have wider application, becoming the foundation or missing link in other research or in the implementation of treatment with personalized medicine methods. In connection with the foregoing, the question of maintaining the confidentiality of personal data when using it in new studies, transferring sets of information elements for the implementation of personalized treatment arises with particular urgency.

Moreover, the cloud computing infrastructure includes various platforms and configurations, as well as a wide variety of security, privacy and authentication settings of different levels and approaches, which creates additional challenges in terms of the security and confidentiality of the relevant data [9]. In addition, earlier in medicine, it was traditionally assumed that privacy can be ensured through the anonymization of personal data (in the case of genomic data, through their generalization), however, the improved methods and algorithms for processing big data and the built connections between various databases put this the concept is in serious doubt. In particular, deanonymization of genetic information from a database is possible using a simple allele frequency, that is, complex DNA can be used to re-identify a person [1].

In this regard, we agree with the researchers who note that one of the optimal solutions to this problem is to minimize the risk of re-identification and protect the right to privacy by placing specific data that allows re-identification in the clouds in predetermined jurisdictions or in special protected areas. including placing critical data outside the cloud [10], as well as the use of virtual computers and sandboxing techniques [9]. It seems that it is necessary to legislatively provide for the authority of the competent federal executive body to regulate the requirements for the placement of data, the use of virtual computers and data management in the cloud, as well as fix the minimum requirements at the level of international regulation.

Separately, we note that the currently accepted model of confidentiality in bioethics provides for informed consent, while from the standpoint of clinical genomic research it is practically impossible to obtain it: firstly, data on health status in open access databases

can be used in many unpredictable ways ; secondly, each time data is loaded, a new object is created, and no specialist or process is responsible for at what point in time or in what way the corresponding data will be available or used [1]; thirdly, the very process of data manipulation, when new genomic classifications are formed, is hidden from the persons providing the data, since the evaluative attitudes used in making appropriate decisions are laid down, inter alia, at the level of a particular laboratory or in software development [11]. In addition, the tendency to include the right to be forgotten in the legislation of various countries also plays a role here. This right is provided for, inter alia, by Regulation 2016/679, which established that the person whose data is used may demand the immediate deletion of his personal data. Corresponding deletion requests can significantly impede the implementation or start of research projects that require the use of a specific dataset, including in the framework of multi-stage surveys, since deleting a single segment of data at one of the stages creates the risk of restarting the entire study [12]. In this regard, we believe that informed consent should cover some generalized options for using the data obtained, in order to ensure a balance between the interests of society (expressed in the normal process of conducting research) and the individual, as well as, in terms of fundamental and applied research, provide for the possibility of waiving the right to be forgotten, since it corresponds to the public interest in the development of science and the application of the corresponding results for the benefit of society and a particular person.

3. ENSURING THE PROTECTION OF THE INTERESTS OF PERSONS

In terms of analyzing the regulation of cross-border transfer of personal data, including genetic data, let us turn to the experience of the European Union. Article 45 of Regulation 2016/679 provides for an "adequate protection" regime - the export of personal data is permitted if the European Commission determines that the privacy laws of the country to which the data are transferred provide adequate protection. The level of protection, among other things, is evidenced by the requirements for data processing activities and for the recipient of information.

Recognition of "adequacy of protection" has been obtained by a small number of countries, including Argentina, Canada, Israel and New Zealand (under the rules of the previously effective European Union Data Protection Directive 95/46 / EC). However, at the same time, another interesting mechanism was established that allows the transfer of information to companies outside the European Union (also within the framework of the invalid Directive 95/46 / EC), namely, in relation

to the United States, whose legislation did not meet the criteria for "adequate protection", the "safe harbor" mode was used. In accordance with this procedure, data could be transferred to specific companies that confirmed their compliance with a number of principles, that is, it was not the state that became the subject of the audit, but a legal entity or a group of legal entities. A third notable option for regulating the transfer of personal data is the Canadian "accountability model", according to which the company that disseminated the data is liable for the actions of the entire chain of persons to whom the data was transferred, based on claims [5].

As a significant private initiative, we also note the framework conditions for the responsible exchange of genomic and health-related data, which were developed under the auspices of the Global Alliance for Genomics and Health and are aimed at creating harmonized approaches to ensure effective and responsible exchange of genomic and clinical data [13].

It seems that the models of "adequate protection", "safe harbor" and "accountability" applied in the complex make it possible to quite effectively ensure the interests of persons whose genetic data is used in cross-border data transfer between research institutes, universities and companies, while the "safe harbor" regime. And framework conditions, agreements, declarations regarding the exchange of genetic data are of particular interest, since they allow certifying specific subjects carrying out the processing and research of genes, including circadian ones.

4. REGULATION OF THE REGIME FOR THE USE OF THE RESULTS OF GENETIC RESEARCH

An important question that arises before participants in scientific research, specialists in the field of genetic engineering and reproductive medicine, commercial structures and other entities is to determine the mode of using the results of genetic research.

So, back in 1996, scientists developed the Bermuda Principles, which stipulate, inter alia, that all data on DNA sequences are not considered as objects of copyright and must be published within twenty-four hours after their appearance, which has formed the modern scientific practice of open access and the concept of treating relevant information as a universal source of knowledge [14].

According to another approach, it is assumed that it is possible to extend patent protection to the results of genetic research. Thus, the US Supreme Court in *Association for Molecular Pathology v. Myriad Genetics*,

Inc. [15] stated that the company was legitimately granted a patent to isolate the BRCA1 and BRCA2 genes in a specific person and to create artificial complementary DNA. The substantiation, among other things, indicated that although the laws of nature, natural phenomena and abstract ideas are the main tools of scientific and technical work and are not subject to patent protection, and the element of DNA that exists in nature belongs to natural objects and is not patentable only because of its isolation, but scientists certainly create new things when they remove elements from DNA to create complementary DNA (such DNA consists only of exons), that is, complementary DNA is patentable, since it does not exist in nature and it can only appear artificial by way. A similar approach is reflected in Directive 98/44 / EC of the European Parliament and the Council of Europe "On the legal protection of biotechnological inventions" of 06.07.1998, since, based on the systemic interpretation of paragraph 24 of the preamble, paragraph 2 of Art. 3, Art. 5, Art. 9 and Art. 23 of this Directive, although a simple discovery of a gene sequence does not form a patentable subject matter, but objects, including naturally formed ones, may be patentable, for which it is necessary, on the one hand, to isolate them from the natural environment, and on the other hand, to describe a specific technical the function of such inventions. The position of researchers who raise the question of the possibility of considering genes as a program is also interesting [16].

At the same time, the research data (manipulations) must satisfy the public interest in terms of observance of the morality generally recognized in a particular country. In this aspect, the practice of the European Court of Human Rights is of interest. Thus, in the judgment in *Parrillo v. Italy* [17], the court considered whether the right to respect for private life under Art. 8 of the Convention for the Protection of Human Rights and Fundamental Freedoms of 04.11.1950, the right to use embryos obtained as a result of in vitro fertilization for the purpose of transferring them to scientific research, and whether the right of ownership extends to them. It was emphasized that from the standpoint of specifically Italian legislation, the human embryo is not considered as a thing, its economic assessment is inadmissible, it is the subject of the right to respect for human dignity. The court concluded that human embryos cannot be reduced to property in terms of Art. 1 of Protocol No. 1 to the European Convention and refused to satisfy the applicant's claims. Also important for the formation of the practice is the case of *Costa and Pavan v. Italy* [18], which positively resolved the issue of the possibility of using assisted reproductive technologies and pre-implantation genetic diagnostics in order to conceive a child not suffering from a genetic disease, provided that the applicants are healthy carriers of such a disease.

Considering the above, we agree with the scientists who note that the civil law regime of genes and genomic technologies is inextricably linked with the right to human life, providing a citizen with qualified medical care, and the process of treating a patient [19]. Thus, further legislative regulation of the use of the results of genetic research, the establishment of copyright, patent protection, and sui generis regimes should be interconnected with the observance of the non-property rights of citizens, taking into account public interests in ensuring moral requirements.

5. CONCLUSIONS

We believe that in order to ensure the right to privacy when using information technologies for the analysis of genetic data, it is necessary to legislatively provide for the authority of the competent federal executive body to regulate the requirements for placing data, using virtual computers and managing data in the cloud, as well as fixing the minimum requirements for the level of international regulation.

With regard to cross-border data transfer, we believe that the complex application of the models of "adequate protection", "safe harbor" and "accountability" allows to effectively ensure the interests of persons whose genetic data is used in the exchange of information between research institutes, universities and companies, while the regime "Safe harbor" and framework conditions, agreements, declarations in terms of the exchange of genetic data are of particular interest, since they allow certifying specific subjects involved in the processing and research of genes, including circadian ones.

The results of genomic research, including in the field of circadian genes, should be used only in compliance with the rights and legitimate interests of the persons whose genetic data is being studied, taking into account ethical and technical aspects.

Ethical issues range from the principled possibility of "patenting life" to the possibility of research or certain manipulation of embryos. At the same time, judicial practice is currently being formed to consider disputes related to research and certain manipulations with genetic material precisely from the position of a balance between public interests (morality) and private non-property interests (respect for private and family life), regardless of the assessment of any property interests. in relation to the relevant objects.

The technical issues of regulating relations in the field of genetics, in particular, are related to the fact that, on the one hand, almost any research involves the isolation of a gene sequence from the natural environment, and, on the other hand, the same gene sequence

can encode the implementation of various processes, in this connection, the question of including in the description of the result of intellectual activity a specific function, which is realized by the corresponding genes, acquires special significance.

Using the example of establishing the legal protection of circadian genes, we believe that in the framework of further discussion it is advisable to consider the development of a comprehensive legal regime for non-property rights to genes (gene sequences) of a person in combination with a sui generis regime of intellectual rights (similar to patent rights or rights to computer programs), while in relation to genes of animals and plants, it is also necessary to take into account public interests (in terms of conducting relevant research).

ACKNOWLEDGMENTS

The reported study was funded by RFBR according to the research project No. 18-29-14073.

REFERENCES

- [1] J. Frizzo-Barker, P. Chow-White, From Patients to Petabytes: Genomic Big Data, Privacy, and Informational Risk, *Canadian Journal of Communication* 39(4) (2014) 615-625. DOI: <https://doi.org/10.22230/cjc.2014v39n4a2743>
- [2] P. Borovska (2017) Big Data Analytics and Genetic Research. <https://www.researchgate.net/publication/323735972> Accessed 11 Mar 2020
- [3] B. Berger, H. Cho (2019) Emerging technologies towards enhancing privacy in genomic data sharing. *Genome Biology*. <https://genomebiology.biomedcentral.com/track/pdf/10.1186/s13059-019-1741-0> Accessed 21 Mar 2020
- [4] J. Adams, Genetics: Big hopes for big data, *Nature* 527(7578) (2015) 108–109. DOI: <https://doi.org/10.1038/527S108a>
- [5] P. Kosseim et al. (2014) Building a data sharing model for global genomic research. *Genome Biology* <https://genomebiology.biomedcentral.com/track/pdf/10.1186/s13059-014-0430-2> Accessed on 22 Feb 2021
- [6] A.A. Inyushkin, Y.S. Povarov, E.S. Krykova, V.D. Ruzanova, Introduction of Genomic Research Results into the Economy Through the Legal Regime of Intellectual Property (by the Example of Circadian Gene Research), in: E. Popkova, B. Sergi (Eds.), *Digital Economy: Complexity and Variety vs. Rationality*, ISC 2019, Lecture Notes in Networks and Systems, vol. 87, Springer, Cham, 2020, pp. 584-590. DOI: https://doi.org/10.1007/978-3-030-29586-8_67
- [7] A.A. Inyushkin, A.N. Inyushkin, V.D. Ruzanova, V.I. Belyakov, E.S. Kryukova, Circadian Genes: Information and Legal Issues, in: E. Popkova, B. Sergi (Eds.), *The 21st Century from the Positions of Modern Science: Intellectual, Digital and Innovative Aspects*, ISC 2019, Lecture Notes in Networks and Systems, vol. 91, Springer, Cham, 2020, pp. 30-37. DOI: https://doi.org/10.1007/978-3-030-32015-7_4
- [8] G. Chassang (2017) The impact of the EU general data protection regulation on scientific research. https://www.researchgate.net/publication/312251732_The_impact_of_the_EU_general_data_protection_regulation_on_scientific_research Accessed 11 Mar 2020
- [9] K.Y. He, Ge. Dongliang, M. He, Research funded by Foundation for the National Institutes of Health, Big Data Analytics for Genomic Medicine, *International Journal of Molecular Sciences* 18(2) (2017) 412. DOI: <https://doi.org/10.3390/ijms18020412>
- [10] F. Molnar-Gabor, R. Lueck, S. Iakhnin, J. O. Korbel (2017) Computing patient data in the cloud: Practical and legal considerations for genetics and genomics research in Europe and internationally. *Genome Medicine*. https://www.researchgate.net/publication/317929677_Computing_patient_data_in_the_cloud_Practical_and_legal_considerations_for_genetics_and_genomics_research_in_Europe_and_internationally Accessed 11 Nov 2020
- [11] P.A. Chow-White, S. Green, Data Mining Difference in the Age of Big Data, *International Journal of Communication* 7(2013) 556-583.
- [12] P. Quinn, L. Quinn (2018) Big genetic data and its big data protection challenges. *Computer Law & Security Review*. https://www.researchgate.net/publication/326411177_Big_genetic_data_and_its_big_data_protection_challenges Accessed 11 Mar 2020.
- [13] B.M. Knoppers, Framework for responsible sharing of genomic and health-related data, *The HUGO journal* 8(1) (2014) 2-6. DOI: <https://doi.org/10.1186/s11568-014-0003-1>
- [14] J. Frizzo-Barker, P.A. Chow-White, A. Charters, D. Ha, Genomic Big Data and Privacy: Challenges and Opportunities for Precision Medicine, *Computer Supported Cooperative Work* 25(2-3) (2016) 115–136. DOI: <https://doi.org/10.1007/s10606-016-9248-7>
- [15] Syllabus and the opinion of the Supreme Court of United States of America on the case Association for Molecular Pathology et al. v. Myriad Genetics, inc.

(2012)

https://www.supremecourt.gov/opinions/12pdf/12-398_1b7d.pdf Accessed on 22 Dec 2020.

[16] A.A. Mohov, A.N. Yavorskiy, Geny I inye obrazovahia na osnove genov kak ob'ekty prava intelektualnoi sobstvennosti, *Gragdanskoe pravo* 4 (2018) 27-31.

[17] Application no. 46470/11 (2015) Parrillo v. Italy, ECHR Judgment of 27 August 2015 <http://hudoc.echr.coe.int/eng?i=001-157263> Accessed on 13 Jan 2021.

[18] Application no. 54270/10 (2012) Costa and Pavan v. Italy, ECHR Judgment of 28 August 2012. <http://hudoc.echr.coe.int/eng?i=001-112993> Accessed 13 Mar 2020

[19] A.N. Levushkin, Civil Law Regime of Genes as Objects of Civil Rights, *Lex russica* 6 (2019) 100-109. DOI: <https://doi.org/10.17803/1729-5920.2019.151.6.100-109>